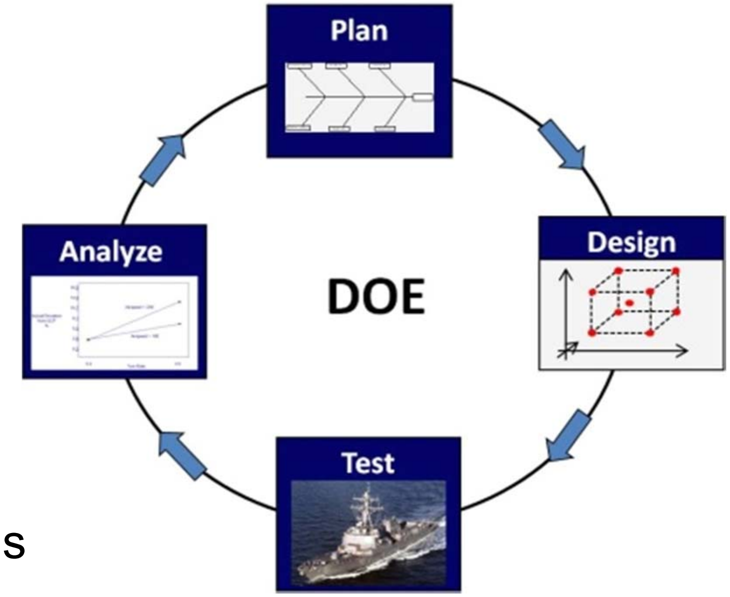
Statistical Analysis Overview

Motivation

- **Design of Experiments**
  - Implemented in test planning stages
  - Programs with successful implementation of DOE are now in the reporting process
  - Appropriate analysis methodologies are not being uniformly applied

- **Objectives:**
  - Provide brief overview of common analysis methodologies
  - Provide conceptual understanding of analysis benefits
  - Provide knowledge of underlying assumptions



All experiments are designed with an analysis methodology in mind – to reap the benefits we need to follow through to the analysis!

Overview

- **Standard Statistical Analysis Methods**
  - Regression
  - Assumptions
- **Model Selection**
- **Advanced Methodologies**
  - Censored Data Analysis
  - Generalized Linear Models
    - » Skewed Data
    - » Logistic Regression
  - Bayesian Methods

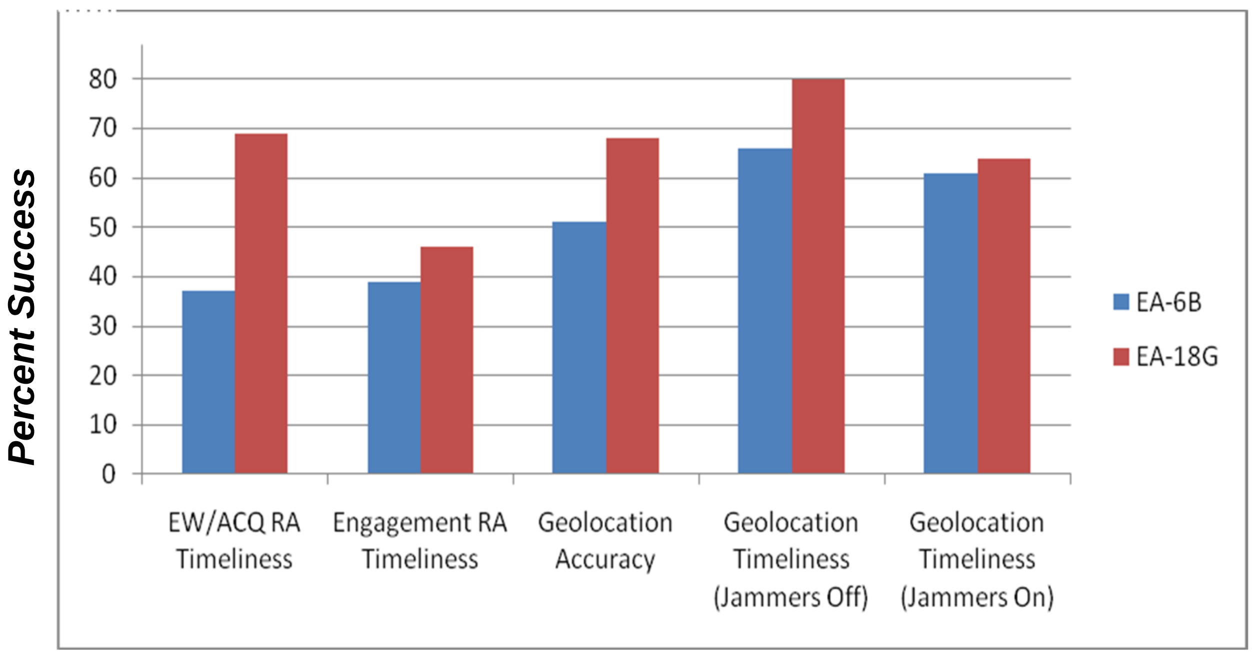
Key Analysis Elements

- **Objective summary of the data:**
  - Quantitative metrics – summarizes what happened in the test
    - » Examples:
      - Percentage of targets detected by target type
      - Median target location error by target type
      - Percentage of messages successfully transmitted
  - Confidence in those results – accuracy of the measurement
    - » Confidence intervals
    - » P-values

- **Determination of Significant Factors**
  - What conditions affect performance? How much?
  - Examples:
    - » The percentage of targets detected is lower for human targets at night than for vehicle targets during the day
    - » Mean detection range is below threshold for cluttered environments, but above threshold for uncluttered environments

EA-18G/EA-6B Comparison Confidence Intervals

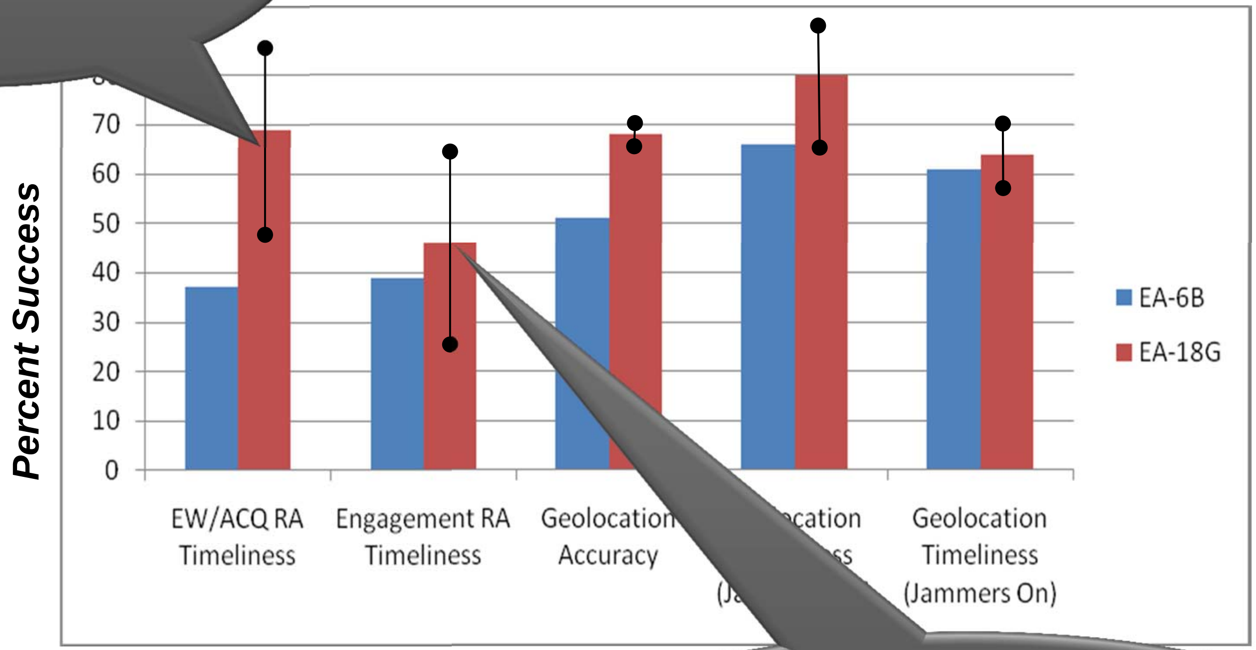
Figure from DOT&E EA-18G BLRIP



EA-18G/EA-6B Comparison Confidence Intervals

from DOT&E EA-18G BLRIP

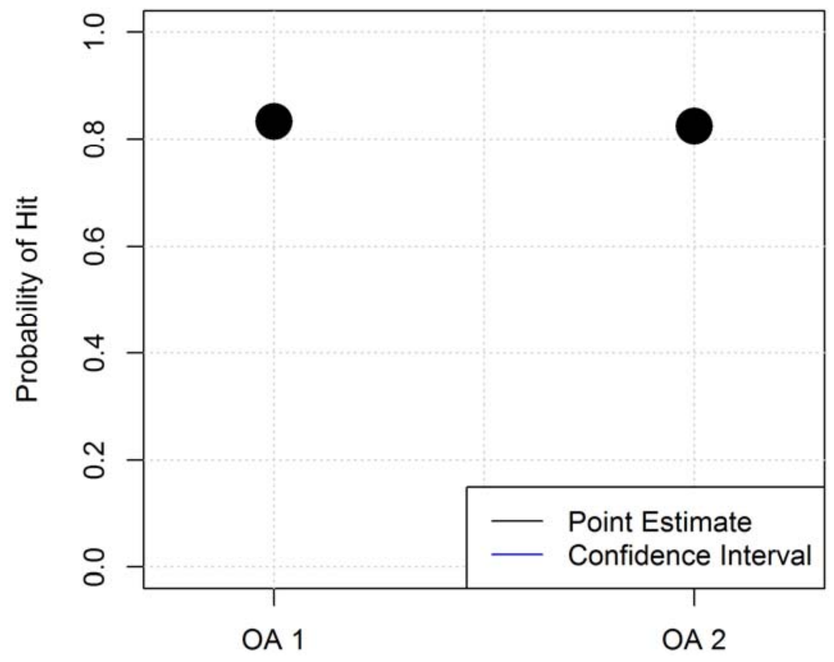
Even without a threshold confidence intervals quantify how accurately the metric was measured



Confidence intervals make it clear that performance is comparable

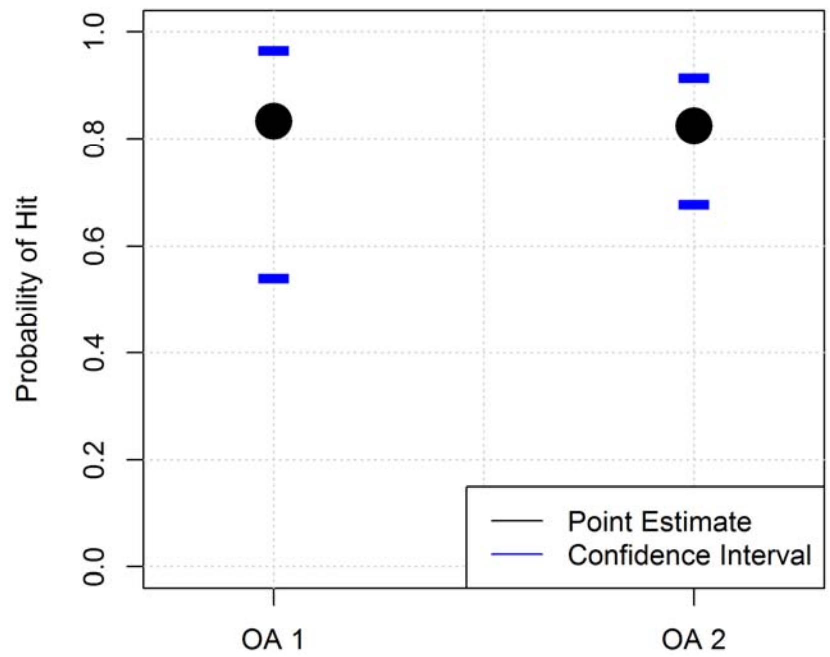
Interval Estimates Are Important

- **Results from a single test event cannot predict exactly how a system will perform in the field**
  - Confidence intervals tell us how precise our test results are
  - More data → tighter confidence bounds
- **Example: New turret for LAV Anti-Tank variant (notional data)**
  - Shoots TOW missiles
  - OA 1: 12 shots
    - » 10 hits
  - OA 2: 40 shots
    - » 33 hits



Interval Estimates Are Important

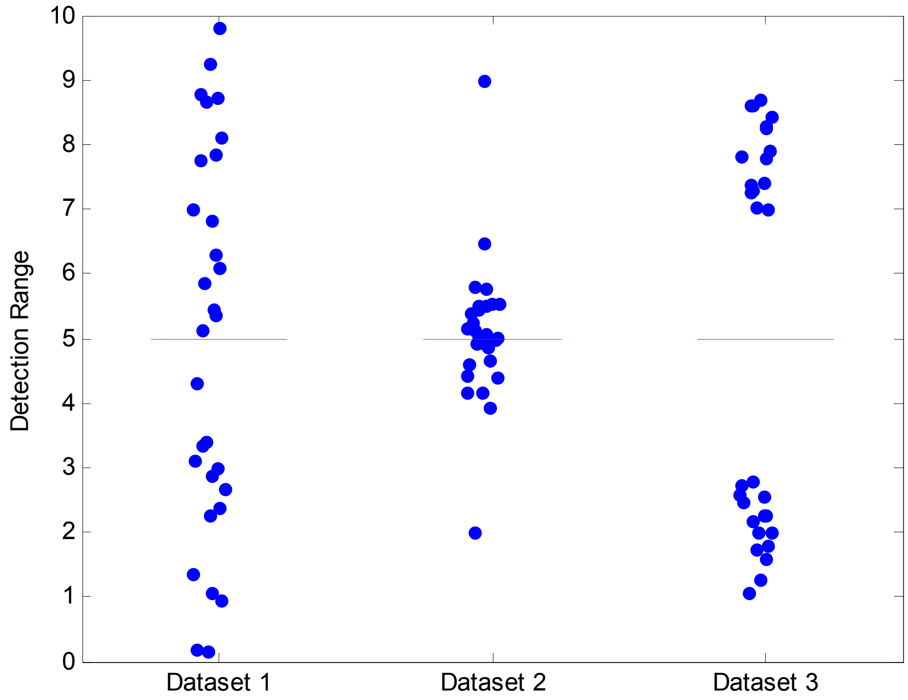
- **Results from a single test event cannot predict exactly how a system will perform in the field**
  - Confidence intervals tell us how precise our test results are
  - More data → tighter confidence bounds
- **Example: New turret for LAV Anti-Tank variant (notional data)**
  - Shoots TOW missiles
  - OA 1: 12 shots
    - » Interval Width: 42.5%
  - OA 2: 40 shots
    - » Interval Width: 23.8%



Interval estimates show the range of values for the system's lifetime performance under similar conditions.

Beware Average Values

- **Using averages (mean) as sole descriptor may miss important information about performance**



Same mean in every case, but very different distributions!

- **The spread (called variance) contains information we want to characterize**
  - Example Dataset 3 – distribution likely due to a significant factor (e.g., different environments or targets)

Overview Parametric Statistical Analyses

Modeling Type		ors	Modeling Assumption
General Linear Model (GLM or LM)		ed	Factors impact mean response through a linear function (on the β's)
Generalized Linear Model	Exponential Family (normal, lognormal, exponential, Poisson,	Fixed	Factors impact one or more model parameters through a linear link function (on the β's)
Linear Mixed Models (LMM)	Normal	Fixed & Random	Factors impact mean response through a linear function (on the β's)
Generalized Linear Mixed Models (GLMM)	Exponential Family	Fixed & Random	Factors impact one or more model parameters through a linear link function (on the β's)
Nonlinear models	Any	Fixed	Any
Nonlinear mixed models	Any	Fixed & Random	Any

Regression
ANOVA
Response Surface Models

Logistic Regression

There are a wide variety of statistical modeling approaches to meet the needs of any data set.

Regression

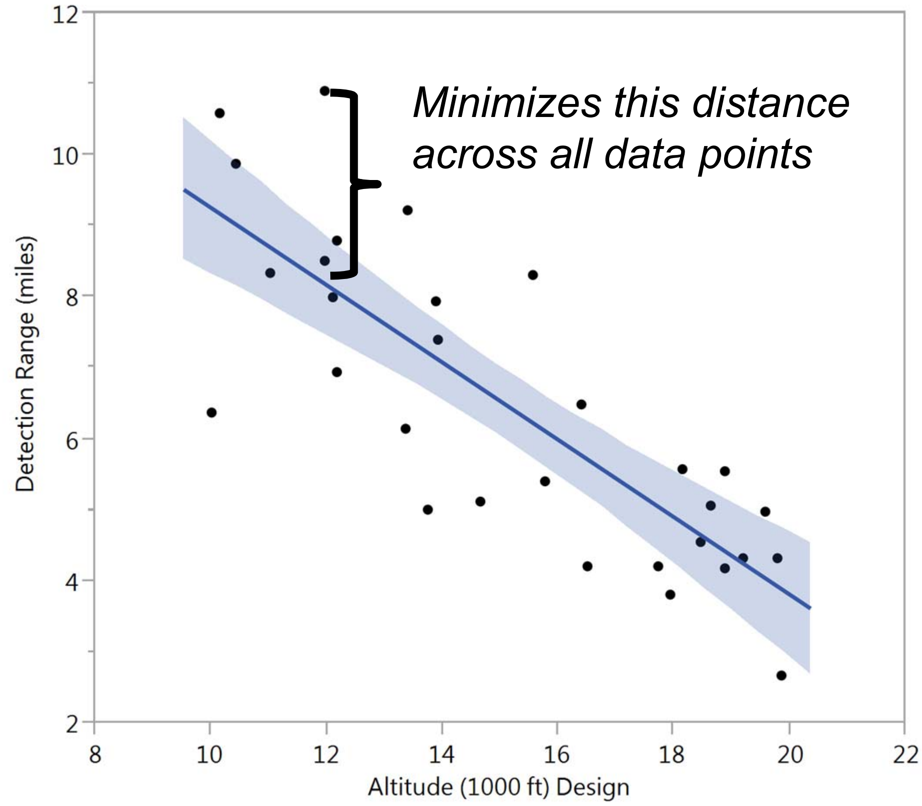
- **Goal: characterize system performance as a function of the predictor variables (factors).**
  - Estimate model parameters (β's)
  - Significant factors are identified by testing hypotheses for model parameters
  - Results:
    - » Confidence Intervals for model parameters
    - » Inferences on mean responses under certain conditions
    - » Predict of future responses
- **Model:**
  - Simple linear regression: $y = \beta_0 + \beta_1 x_1 + \varepsilon$
  - Multiple linear regression: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$
- **Hypothesis test on the slope:**
  - H_0: $\beta_1 = 0$ (no effect)
  - H_1: $\beta_1 \neq 0$ (effect of factor non-zero)

Simple Linear Regression

- **Intelligence, Surveillance, and Reconnaissance (ISR) notional example**

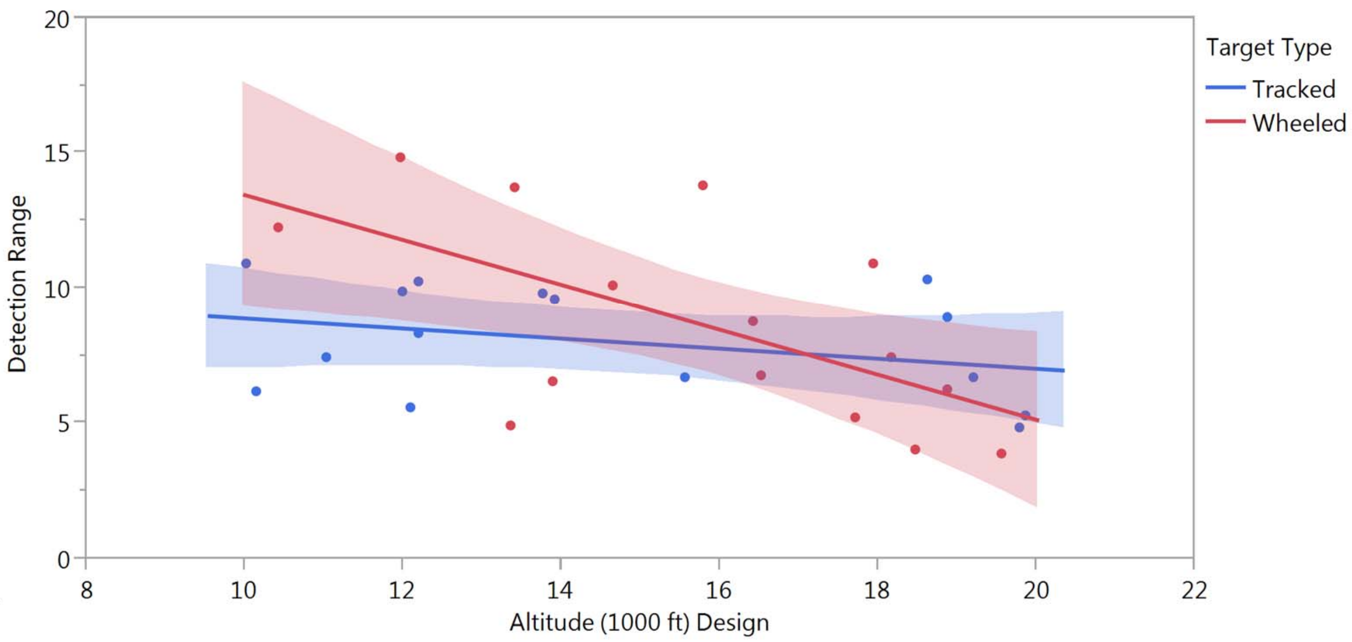
- **Regression (least squares)**
  – Minimizes discrepancy (squared distance) between data and the regression line
  – Here β_0 is the intercept and β_1 is the slope

- **Maximum Likelihood Estimation (MLE)**
  – Finds the most likely value of the parameters (β_0, β_1) given the observed data, by maximizing the likelihood function



Extension to Multiple Linear Regression

- **Two types of targets, wheeled and tracked vehicles**
- **Data recoding – can make any analysis a regression**
  - Any categorical factor with k levels can be recoded as a k-1 indicator variables
  - For example: $x_2 = \begin{cases} 1, if\ wheeled \\ 0, if\ tracked \end{cases}$
- **The result in this simple analysis is essentially two regression equations**
  - We had to include the two-factor interaction to see the difference between tracked and wheeled vehicles. This is why two-factor interactions are important!



Regression Assumptions

1. **Independence: Responses are independent of one another**

2. **Normality: Each population of responses follow a normal distribution**

3. **Homoscedasticity: Each normal distribution has a common variance**

4. **Linearity: Linear relationship between overall mean and the coefficients of the model terms**

The goal of regression analysis is not to correctly model the underlying physics, rather it is to explain as much variation in the data as possible!

In other words, we know the model is wrong, but it may be useful.

We must check the assumptions to understand if the model is useful!

Regression Example

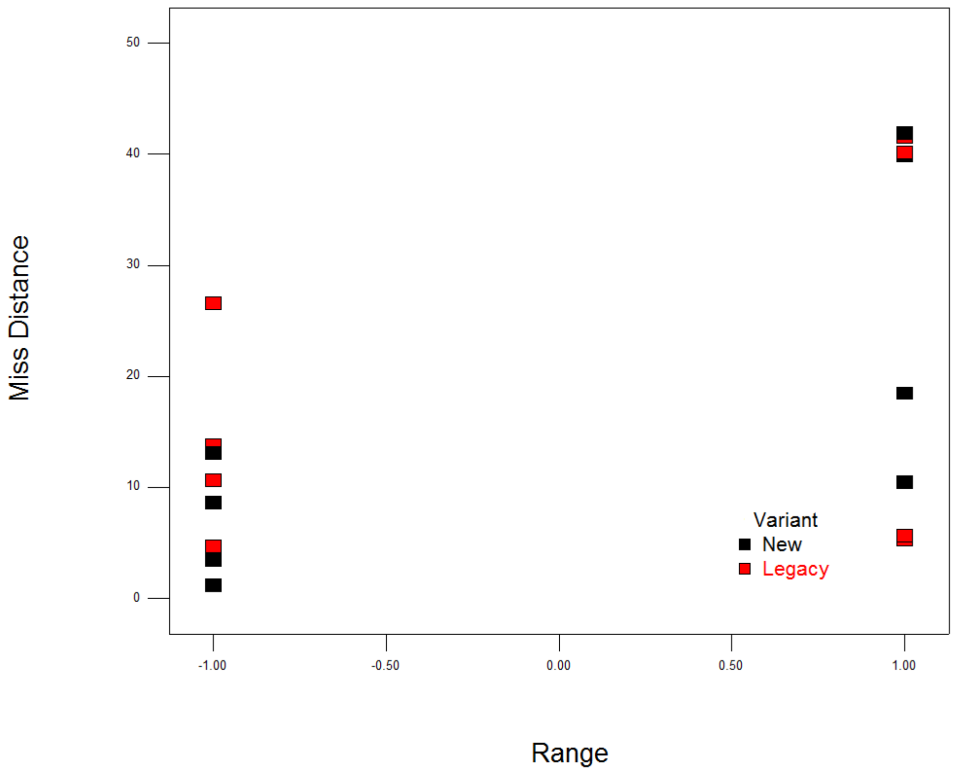
- **Air to Ground Missile Test**
  - Objectives:
    1. Characterize performance of a new air-to-ground missile
    2. Compare the new missile to legacy
  - Response variable: miss distance
  - Factors: range to target, altitude, speed, variant (new versus legacy)

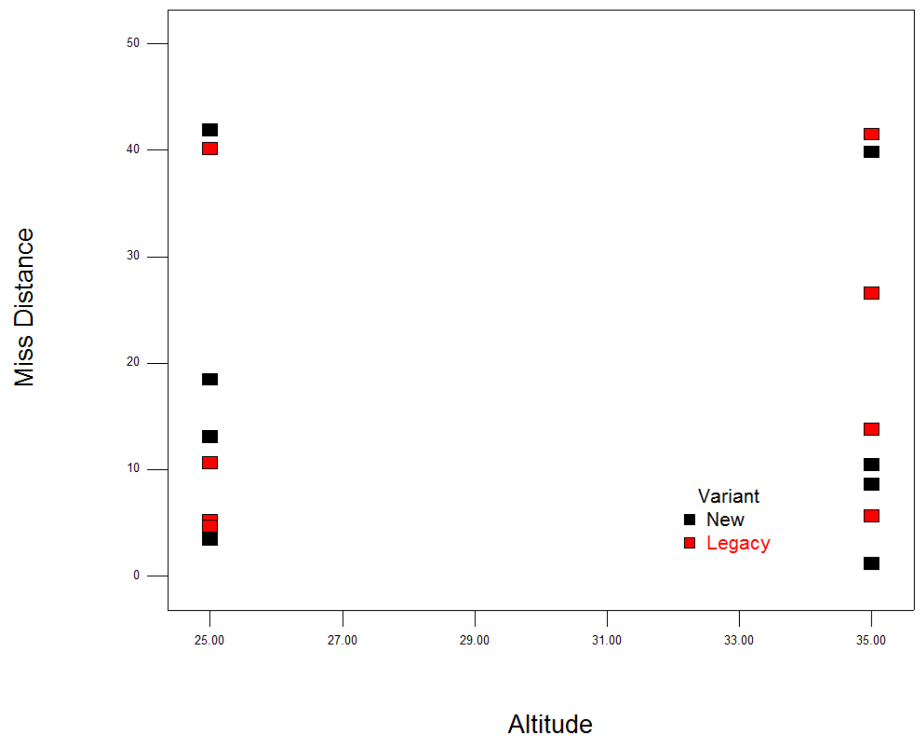
- **Test Design**
  - Full factorial, 16 run screening design

Run	Variant (coded)	Range	Altitude	Airspeed	Miss Distance
1	New (1)	-1	35	0.85	1.14
2	Legacy (0)	1	35	0.95	41.47
3	New (1)	1	25	0.85	18.45
4	Legacy (0)	-1	35	0.95	13.76
5	New (1)	1	35	0.95	39.81
6	Legacy (0)	1	25	0.85	5.23
7	New (1)	-1	25	0.85	13.04
8	Legacy (0)	1	35	0.85	5.63
9	New (1)	1	25	0.95	41.90
10	New (1)	-1	35	0.95	8.58
11	Legacy (0)	1	25	0.95	40.09
12	Legacy (0)	-1	25	0.85	4.65
13	Legacy (0)	-1	35	0.85	26.55
14	Legacy (0)	-1	25	0.95	10.58
15	New (1)	1	35	0.85	10.44
16	New (1)	-1	25	0.95	3.44

Exploratory Data Analysis

- **Provides intuition on best analysis model**

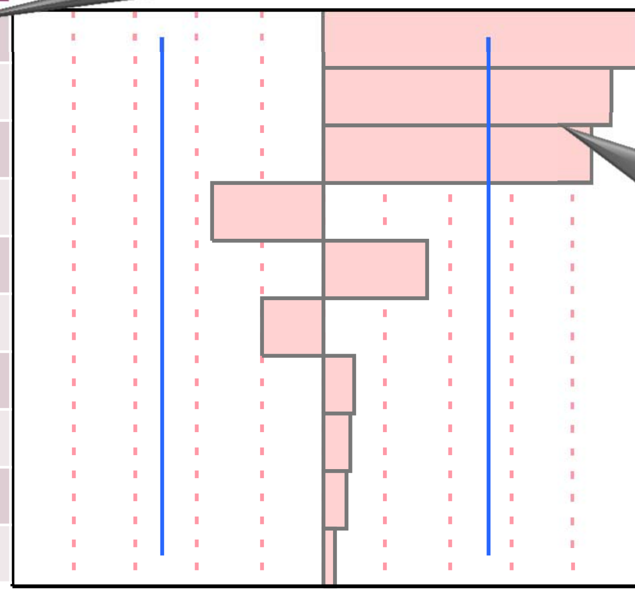




IDA

Important Factors

Model Term	P-Value
range*air speed	**0.0043***
range	**0.0062***
air speed	**0.0079***
variant*range	0.1363
variant*altitude	0.1636
range*altitude	0.3657
variant*air speed	0.6436
variant	0.6943
altitude	0.7242
altitude*air speed	0.8532



Small p-value means there's little chance the change in performance when changing this factor is due to chance alone

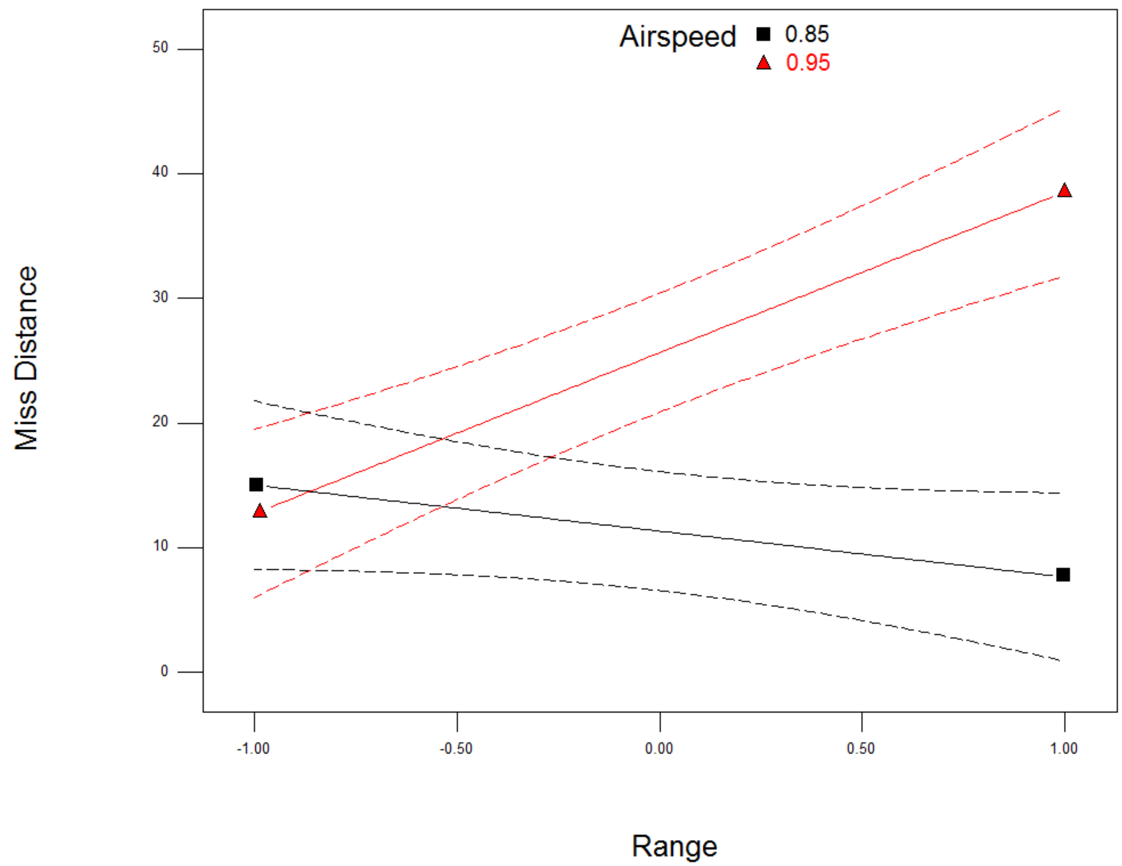
This chart shows us how significant a factor is on performance relative to the noise in the data

- **<u>Conclusion:</u> Range and airspeed are the two most important factors in characterizing performance for both the new and legacy air to ground missiles**

- **On average, there is no statistically distinguishable difference between the two variants across the operational envelope investigated in this test**

Graphical Presentation of Results

- **Interaction plots provide meaningful insights**
  - Miss distance only increases with range at the higher velocity



IDA

Analysis Key Takeaways

- **We use statistics and robust analysis to ensure we have defensible conclusions**
  - Confidence intervals tell us how accurately we measured the KPP/MOE, how confident we are in claiming it met requirements
- **DOE methodology ensures we *characterize performance* across the operational envelope**
  - Avoid the average value, which hides important factors
- **There are lots of analysis tools available**
  - Analysis methods provide means for data visualization
  - Statistical analysis methods enable more precise measurement/knowledge of system performance (more with less)
- **What's next:**
  - Statistical Model Selection
  - Advanced Techniques